OPIC
OFFICE DE LA PROPRIÉTÉ INTELLECTUELLE DU CANADA

CIPO
CANADIAN INTELLECTUAL PROPERTY OFFICE

Ottawa Hull K1A 0C9

(21) (A1) 2,170,611
(86) 1994/09/02
(43) 1995/03/09

(51) Int.Cl.[6] C12N 15/53; C12N 15/11; C12N 15/82; C12N 15/87; C12N 5/10; A01H 5/00

(19)(CA) **APPLICATION FOR CANADIAN PATENT** (12)

(54) Glycerin-3-Phosphate-Dehydrogenase (GPDH)

(72) Töpfer, Reinhard - Germany (Federal Republic of) ;
Hausmann, Lüdger - Germany (Federal Republic of) ;
Schell, Jozef - Germany (Federal Republic of) ;

(71) Max-Planck-Gesellschaft zur Förderung der Wissenschaften e.V. - Germany (Federal Republic of) ;

(30) (DE) P 43 29 827.3 1993/09/03

(57) 12 Claims

MAX PLANCK SOCIETY

for Promotion of the Sciences e.V. [Registered Association]

37073 Goettingen

Glycerol-3-phosphate dehydrogenase (GPDH)

This invention concerns DNA sequences that code for a glycerol-3-phosphate dehydrogenase (GPDH) and the alleles as well as the derivatives of these DNA sequences.

This invention also concerns genomic clones that contain the complete gene of a glycerol-3-phosphate dehydrogenase and alleles as well as derivatives of this gene.

This invention also concerns promoters and other regulator elements of glycerol-3-phosphate dehydrogenase genes.

Glycerol-3-phosphate dehydrogenase (GPDH; EC 1.1.1.8), also known as dihydroxyacetone phosphate reductase, is substantially involved in triacylglyceride biosynthesis in plants by supplying glycerol-3-phosphate. Fatty acid biosynthesis and triacylglyceride biosynthesis can be regarded as separate biosynthesis pathways owing to compartmentalization but as one biosynthesis pathway from the standpoint of the end product. *De novo* biosynthesis of fatty acids takes place in the plastids and is catalyzed by three enzymes or enzyme systems, i.e., (1) acetyl-CoA carboxylase (ACCase), (2) fatty acid synthase (FAS), and (3) acyl-[ACP]-thioesterase (TE). The end products of this reaction sequence in most organisms are either palmitic acid, stearic acid, or after desaturation, oleic acid.

In the cytoplasm, however, triacylglyceride biosynthesis takes place via the so-called "Kennedy pathway" in the endoplasmic reticulum from glycerol-3-phosphate which is made available by the activity of glycerol-3-phosphate dehydrogenase (S.A. Finnlayson et al., Arch. Biochem. Biophys., 199 (1980)

pages 179-185), and from fatty acids present in the form of acyl-CoA substrates.

Probably the first discovery of the enzymatic activity of glycerol-3-phosphate dehydrogenase in plants involved potato tubers (G.T. Santora et al., Arch. Biochem. Biophys., 196 (1979) pages 403-411). This activity had not been observed in other plants before then (B. König and E. Heinz, Planta, 118 (1974) pages 159-169), so the existence of the enzyme had not been detected. Thus the formation of glycerol-3-phosphate on the basis of the activity of a glycerol kinase was discussed as an alternative biosynthesis pathway. Santora et al., loc. cit., subsequently detected GPDH in spinach leaves and succeeded in increasing the concentration of the enzyme approximately 10,000 times. They determined the native molecular weight to be 63.5 kDa and found the optimum pH for the reduction of dihydroxyacetone phosphate (DHAP) to be 6.8 to 9.5 for the back reaction. GPDH was likewise detected in Ricinus endosperm (Finlayson et al., Biochem. Biophys. 199 (1980) pages 179-185). According to more recent works (Gee et al., Plant Physiol. 86 (1988a) pages 98-103), two GPDH activities could be detected in enriched fractions, a cytoplasmic fraction (20-25%) and a plastid (75-80%). The two forms are regulated differently. Thus, for example, the cytoplasmic isoform can be activated by F2,6DP, while the plastid isoform is activated by thioredoxin (R.W. Gee et al., Plant Physiol., 86 (1988) pages 98-103 and R.W. Gee et al., Plant Physiol., 87 (1988) pages 379-383).

The methods of molecular biology are making increasing entry into plant cultivation practice. Changes in biosynthesis output with the formation of new components and/or higher yields of these components can be achieved with the help of gene manipulation, e.g., transfer of genes which code for enzymes. As one of the most important enzymes of triacylglyceride synthesis, GPDH has a significant influence on the oil yield of plants.

It is thus the object of this invention to improve the oil yield of crop plants by influencing the triacylglyceride content.

This object is achieved with the DNA sequences according to patent claim 1 and the genes from the genomic clones according to patent claim 4.

This invention concerns DNA sequences that code for a glycerol-3-phosphate dehydrogenase, and alleles as well as derivatives of these DNA sequences.

This invention also concerns genomic clones that contain a complete gene of a glycerol-3-phosphate dehydrogenase including the structure gene, the promoter and other regulator sequences, and alleles as well as derivatives of this gene.

This invention likewise concerns the promoters and other regulator elements of glycerol-3-phosphate dehydrogenase genes from the specified genomic clones, and the alleles as well as derivatives of these promoters.

This invention additionally concerns a method of producing plants, plant parts and plant products in which the triacylglyceride content or fatty acid content is altered, where DNA sequences or genes are transferred from the genomic clones by the methods of genetic engineering.

This invention also concerns the use of said DNA sequences or one of the genes originating from said genomic clones for altering the triacylglyceride content or its fatty acid pattern in plants.

Finally, this invention concerns transgeneic plants, plant parts and plant products produced according to the aforementioned method.

The figures serve to clarify the present invention.

They show the following:

Figure 1: Comparison of the derived amino acid sequences of the ClGPDH30 and CLGPDH109 cDNAs as well as the gene from the ClGPDHg3 genomic clone with the GPDH amino acid sequence of the mouse (Mm GPDH);

Figure 2: Separation of proteins from BB26-36 cells by gel electrophoresis;

Figure 3: Map of the insertions contained in ClGPDHg5, ClGPDHg9 and ClGPDHg3 genomic clones with various restriction enzymes;

Figure 4: Schematic diagram of the functional areas of the genes contained in the ClGPDH5, ClGPDH9 and ClGPDH3 genomic clones; and

Figure 5: Northern Blot with RNAs from various plant tissues, hybridized with ClGPDH20 cDNA as a probe.

It is obvious that allelic variants and derivatives of DNA sequences or genes according to this invention are included within the scope of this invention under the assumption that these modified DNA sequences or modified genes will code for glycerol-3-phosphate dehydrogenase. The allelic variants and derivatives include, for example, deletions, substitutions, insertions, inversions and additions to DNA sequences or genes according to this invention.

Any plant material that produces glycerol-3-phosphate dehydrogenase in sufficient quantities is a suitable raw material for isolating cDNAs that code for glycerol-3-phosphate dehydrogenase. Isolated embryos from the plant Cuphea lanceolata, indigenous to Central America, have proven to be an especially suitable raw material in the present invention.

Functional complementation was used for isolation of DNA sequences according to this invention. This refers to complementation of mutant microorganisms with heterologous cDNA. Functional complementation was performed after infecting E. coli strain BB26-36, which is auxotrophic for glycerol, with phagemids containing plasmids with cDNAs from Cuphea lanceolata. Plasmids isolated from functionally complemented bacteria were cleaved with

restriction endonucleases and separated by electrophoresis. The cDNAs contained in the plasmids were classified in two classes that differ in the size of their insertions. Retransformation confirmed that the isolated cDNAs were capable of complementing the BB26-36 mutant.

The complete coding area of one of the two classes codes for a glycerol-3-phosphate dehydrogenase and is contained in the ClGPDH20 cDNA clone. This is an Eco RI-ApaI fragment that has 1354 base pairs. The complete 1354 base pair DNA sequence of the ClGPDH20 cDNA and the amino acid sequence derived from it are entered in the Sequence Listing as SEQ ID NO:1. ClGPDH20 cDNA was sequenced double stranded. Proceeding from the ATG start codon, the cDNA codes from positions 17 to 1132 for a protein with 372 amino acids (ending at the TAG stop codon), which is expressed as a fusion with lacZ without a shift in the reading frame. The estimated molecular weight is 40.8 kDa. Two base pairs (CA) preceding ATG are included with the cDNA. The first 14 nucleotides are attributed to the DNA sequence of the fusion with lacZ, and the linker sequence is indicated at the 3' end. The polyA signal is found at positions 1329 to 1334 in the 3' untranslated region.

It is assumed that ClGPDH20 cDNA is a cytoplasmic isoform, because no transit peptide can be detected in homology comparisons with mouse GPDH (see Figure 1). On the basis of the position of an assumed NADH binding site corresponding to the consensus sequence GxGxxG (see positions 29 to 34 in the ClGPDH20 amino acid sequence in Figure 1 (R.K. Wierenga et al., Biochem. 24 (1985) pages 1346-1357), the N-terminal sequence of 28 amino acids is not sufficient to code for a transit peptide whose length varies between 32 and 75 amino acids (Y. Gavel et al., FEBS Lett. 261 (1990) pages 455-458).

A cDNA library from Cuphea lanceolata was screened with ClGPDH20 cDNA as a probe for isolation of additional GPDH cDNAs, and a total of 52 cDNA clones

were isolated. The 18 longest cDNAs were completely or partially sequenced. The ClGPDH109, ClGPDH30 and ClGPDH132 cDNA clones contain cDNAs with the complete coding region or a virtually complete cDNA of GPDH.

The ClGPDH109 cDNA clone contains the complete coding region of GPDH on a 1464 base pair EcoRI-ApaI DNA fragment which codes for a protein with 381 amino acids. The DNA sequence and the amino acid sequence derived from it are shown as SEQ ID NO:2 in the Sequence Listing. The DNA fragment was sequenced double stranded. The coding area begins with the ATG start codon in position 45 and ends in position 1187, followed by the TAG stop codon (positions 1188 to 1190). The cDNA itself begins at position 15. The first 14 nucleotides are attributed to the DNA sequence of the fusion with lacZ. The polyA signal (positions 1414 to 1419) and the polyA area (positions 1446 to 1454) as well as the linker sequence (positions 1459 to 1464) are found in the untranslated region at the 3' end.

Another cDNA, ClGPDH30, also contains the complete coding region of GPDH on a 1390 base pair EcoRI-XhoI fragment, which codes for a protein with 372 amino acids. The double-stranded-sequenced DNA sequence and the DNA sequence derived from it are listed as SEQ ID NO:4 in the Sequence Listing. The protein coding sequence begins with the ATG start codon at position 34 and ends before the stop codon at position 1149. The first 14 base pairs are attributed to the sequence of the fusion with lacZ. The polyA signal (positions 1349 to 1354) and the polyA region (positions 1366 to 1384) are found in the untranslated 3' area.

The ClGPDH132 cDNA clone with 1490 base pairs is an Eco RI-XhoI fragment, the DNA sequence of which and the amino acid sequence derived from it are shown as SEQ ID NO:3 in the Sequence Listing. The DNA fragment was sequenced double stranded. ClGPDH132 cDNA is missing 14 amino acids at the N terminus In

comparison with ClGPDH109 cDNA. The open reading frame begins at position 15 and ends at position 1115, followed by the stop codon at positions 1116 to 1118. Consequently, ClGPDH132 cDNA codes for a protein with 367 amino acids and likewise includes the coding area for glycerol-3-phosphate dehydrogenase with the exception of 14 amino acids. The first 14 nucleotides are to be attributed to the lacZ fusion sequence and the linker sequence (positions 1485 to 1490) is at the 3' end. The polyA signal and the polyA area are located at positions 1343 to 1348 and 1465 to 1484, respectively, in the untranslated 3' area.

Two classes of cDNAs can be distinguished on the basis of sequence data. Accordingly, ClGPDH20 and ClGPDH30 cDNAs belong to class A and ClGPDH132 and ClGPDH109 cDNAs belong to class B.

As Figure 1 shows, the derived amino acid sequences of ClGPDH30 and ClGPDH109 cDNAs show 96% identical amino acids. At the same time, the derivative amino acid sequences of the cDNAs and those of a gene to be assigned to another class, ClGPDH30, were compared with the GPDH amino acid sequence of the mouse (MmGPDH). The differences between the amino acid sequence derived from the ClCPDH109 cDNA, the coded amino acid sequence of the gene and the mouse GPDH in comparison with the amino acid sequence derived from ClGPDH30 are shown in black. On the average, the identity of the derivative proteins of the cDNAs and the GPDH gen with the mouse protein is approximately 50%.

ClGPDH20 cDNA was cloned into an expression vector and expressed in E. coli as a fusion protein with glutathione-S-transferase. To do so, the cDNA was cloned beginning with ATG (see position 17, SEQ ID NO:1) into pGX, a derivative of the pGEXKG expression vector (K.L. Guan et al., Analytical Biochem. 192 (1991) pages 262-267). BB26-36 cells were harvested at various times after administration of IPTG (isopropyl-b-thiogalcatopyranoside) and

their proteins were separated by gel electrophoresis. Figure 2 shows gel electrophoretic separation of BB26-36 cell extracts. The left column shows the proteins of cells with the pGX expression vector (without fusion; 26 kDa protein) and the right side shows proteins of cells with the pGXGPDH20 expression vector which codes for a fusion protein of 67 kDa. The hourly values given indicate the times of sampling after IPTG induction. This clearly shows an enrichment of the fusion protein after two hours. An enzyme activity determination was subsequently performed by enzyme assay of GPDH with an isolated fusion protein and significant enzyme activity was measured. This finding clearly proves that ClGPDH20 cDNA contains a competent gene for expression of GPDH.

Furthermore, genomic clones were isolated, where a library of genomic DNA of **Cuphea lanceolata** was screened with ClCPDH20 cDNA as a probe. By this method, 31 genomic clones were isolated. The genomic clones contain a complete structure gene of a glycerol-3-phosphate dehydrogenase and alleles plus derivatives of this gene together with the promoter sequence and other regulator elements. This means that they form complete transcription units.

Three genomic clones are characterized below. These include the ClGPDHg3 genomic clone with a 15.9 kb DNA insertion, the ClGPDHg5 genomic clone with a 17.7 kb DNA insertion, and the ClGPDHg9 genomic clone with a 15.6 kb DNA insertion. Figure 3 shows a map of the DNA insertions of the genomic clones with various restriction enzymes. The black bars indicate the fragments that hybridize with a 5' probe of the GPDH20 cDNA. The white bars show the areas of DNA insertions that were sequenced and are included in the Sequence Listing.

Sequence analysis of the areas presented in Figure 3 (white bars) of the three genomic clones ClGPDHg5, ClGPDHg3 and ClGPDHg9 has shown that they

contain the complete or partial structure gene of GPDH with all or most of the promoter sequence (5' direction). Figure 4 shows a schematic diagram of the sequenced areas of the genomic clones. The ClGPDHg5, ClGPDHg9 and ClGPDHg3 genomic clones contain the complete structure genes of GPDH in addition to promoter sequences. The entire promoter of GPDH was sequenced from the ClGPDHg9 genomic clone.

Thus a 4434 bp DNA fragment of the ClGPDHg5 genomic clone contains parts of the promoter and the complete structure gene of GPDH in the 5' area. The double-stranded-sequenced DNA sequence as well as the amino acid sequence derived from it are shown as SEQ ID NO:5 in the Sequence Listing. The protein-coding sequence interrupted by DNA areas not translated (introns) with 372 amino acids begins with the ATG start codon in position 1394 and ends before the TAG stop codon in position 4005. The putative TATA box is located at positions 1332 to 1336. Transcription presumably starts at position 1364 (Joshi, NAR 15 (1987) pages 6643-6653). The polyA signal is located in positions 4205 to 4210 at the 3' end. Position 4221 corresponds to the last nucleotide before the polyA area of ClGPDH30 cDNA (see position 1365 in SEQ ID NO:4).

The complete structure gene of GPDH as well as parts of the promoter in 5' direction are contained in a 4006 bp DNA fragment from the ClGPDHg3 genomic clone. The DNA sequence of the DNA fragment that was sequenced mostly as a double strand from ClGPDHg3 as well as the amino acid sequence derived from it are shown as SEQ ID NO:6a and SEQ ID NO:6b in the Sequence Listing. The protein coding area interrupted by intron sequences begins at position 1182 (see SEQ ID NO:6a) with the ATG start codon and ends with the TAG stop codon at position 190 (see SEQ ID NO:6b). CAAT box and TATA box signal sequences are located at positions 1055 to 1058 and 1103-1107 before the start of

transcription. Assumed transcription starting points are at positions 1136 and 1148. Owing to a lack of sequence data, an area of approximately 480 base pairs is not identified within the coding sequence. The polyA signal is located in the untranslated 3' area at positions 393 to 398 (SEQ ID NO:6b).

The entire promoter as well as the first exon of the sequence coding for GPDH are contained in a 1507 bp DNA fragment from the ClGPDHg9 genomic clone. The DNA sequence that was sequenced mostly as a double strand as well as the amino acid sequence derived from it are shown as SEQ ID NO:7 in the Sequence Listing. The TATA box is located at positions 1108 to 1112 before the start of transcription. The protein coding sequence begins with the ATG start codon at position 1193 and ends at position 1376, where an untranslated area (intron) begins. Transcription presumably starts at position 1144.

By comparing DNA sequences, it has been found that ClGPDH30 cDNA, which includes a complete protein reading frame for GPDH, is identical to the GPDH gene from the ClGPDHg5 genomic clone. Consequently, the ClGPDHg5 genomic clone can be classified in class A (see above). The ClGPDH132 cDNA with an almost complete protein reading frame for GPDH is identical to the gene from the ClGPDHg9 genomic clone, which consequently may be assigned to class B (see above). The gene from the ClGPDHg3 genomic clone cannot be assigned to either of the two classes, and thus forms another class C.

Genetic engineering methods (in the form of anti-sense expression or overexpression) can be used to introduce or transfer the DNA sequences according to this invention that code for a glycerol-3-phosphate dehydrogenase into plants for the production of these dehydrogenases for the purpose of altering the biosynthesis yield of these plants. Inasmuch as the DNA sequences according to this invention are not a complete transcription unit, they are preferably introduced into the plants together with suitable promoters,

especially in recombinant vectors, such as binary vectors. Genomic clones can be used as separate complete transcription units for the transformation of plants in order to influence the triacylglyceride content and the fatty acid distribution.

Any species of plants can be transformed for this purpose. Oil-bearing plants, such as rapeseed, sunflower, linseed, oil palm and soybean are preferred for this transformation in order to influence the triacylglyceride biosynthesis in these plants in the manner desired.

The introduction of DNA sequences according to this invention that code for a glycerol-3-phosphate dehydrogenase as well as the complete genes contained in the genomic clones of a glycerol-3-phosphate dehydrogenase by the methods of genetic engineering can be performed with the aid of conventional transformation techniques. Such techniques include direct gene transfer, such as microinjection, electroporation, use of particle gun, steeping plant parts in DNA solutions, pollen or pollen tube transformation, viral vector-mediated transfer and liposome-mediated transfer as well as the transfer of appropriate recombinant Ti plasmids or Ri plasmids through **Agrobacterium tumefaciens** and transformation by plant viruses.

The DNA sequences according to this invention as well as the complete genes of a glycerol-3-phosphate dehydrogenase contained in the genomic clones are excellent for achieving a significant increase in oil production by transgeneic plants. This increase in oil yield is obtained with an increase in triacylglyceride content in of the seed due to overexpression of GPDH. Furthermore, a reduction in glycerol-3-phosphate dehydrogenase can be obtained through anti-sense expression or cosuppression, so the building blocks for triacylglyceride synthesis are missing. This effect is especially beneficial when the production of wax esters (such as jojoba wax esters) in the seeds of

transgeneic plants is to be improved. Another possible application of DNA sequences according to this invention as well as the genes from the genomic clones would be for suppressing triacylglyceride biosynthesis in transgeneic plants and making available the CoA ester as well as glycerol-3-phosphate for other biosyntheses.

Moreover, the promoters of glycerol-3-phosphate dehydrogenase genes from clones according to this invention can, for example, be used for targeted expression of chimeric genes in embryo-specific tissue. On the basis of experimental data it is assumed with regard to the specificity of the promoters that the promoters of genes from the ClGPDHg5 and ClGPDHg9 genomic clones are seed-specific, while the promoter of the gene from the ClGPDHg3 genomic clone has little or no activity in the embryo. Thus, for example, a 1387 bp BamHI/AlwNI fragment of ClGPDHg5 is suitable for transcriptional fusion, a 1189 base pair SphI/NarI fragment of ClGPDHg9 is suitable for translational fusion and a 1172 base pair BamHI/BsmAI (part.) fragment of ClGPDHg3 is suitable for transcriptional fusion. Larger (or smaller) promoter fragments can be used for expression of chimeric genes on the basis of additional clones present on the genetic clones. Likewise, any regulatory sequences located downstream from the first codon of the GPDH gene are obtained for targeted expression of chimeric genes from the cloned fragments of genomic DNA.

Northern Blot analysis with polyA$^+$-RNA from various Cuphea lanceolata tissues with ClGPDH20 cDNA as a probe shows very large amounts of RNA in embryos in comparison with other tissues (see Figure 5). The increase in RNA correlates with increased gene expression and consequently indicates an extremely strong promoter.

The following examples are presented to illustrate this invention.

EXAMPLES

The plant material used in the context of the present invention was obtained from **Cuphea lanceolata** (Lythraceae) (small lanceolate tube flower).

Example 1

Production of glycerol-3-phosphate dehydrogenase cDNAs from Cuphea lanceolata

A cDNA library was prepared from **Cuphea lanceolate** (wild type) took place with the help of the ZAP® cDNA synthesis kit according to the manufacturer's instructions (Stratagene, La Jolla, USA). Messenger RNA from isolated immature embryos about two to three weeks old was used as raw material for the synthesis of the cDNAs. The cDNA library obtained in this way contained 9.5×10^5 recombinant phages.

Functional complementation for isolation of cDNAs that code for a glycerol-3-phosphate dehydrogenase was performed with the E. Coli BB26-36 strain (R.M. Bell, J. Bact. 117 (1974) pages 1065-1076). The bacterial medium for culturing BB26-36 (bearing the plsB26 and plsX mutations) was supplemented with 0.1% glycerol to supplement the bacteria. A medium without glycerol was used for functional complementation.

The pBluescript plasmids were cut out of the above cDNA library in l-ZAP II according to the manufacturer's instructions (Stratagene) by in vivo excision using helper phages and then packed in phage coats: 200 ml of XL1Blue E. Coli cells ($OD_{600} = 1$) were infected with 5×10^5 pfu of the l-ZAP II cDNA library, and, in order to guarantee coinfection, were also infected with a tenfold amount of f1 R408 helper phages. After incubating for 15 minutes at a temperature of 37°C for phage adsorption, 5 ml 2xYT medium were added and agitated for three hours more at a temperature of 37°C. During this time, the cells of the pBluescript plasmids packed in the coats of helper phages are secreting the so-called phagemids into the medium. The bacteria were killed

and the l phages were inactivated by a heating for 20 minutes at 70°C. After centrifuging, the supernatant containing helper phages along with phagemids was removed. This supernatant was used for infection of the mutant BB26-36 strain.

Complementation was performed after infecting the E. coli BB26-36 strain with phagemids containing cDNA plasmids that code for a glycerol-3-phosphate dehydrogenase. M56-LP medium (Bell, *loc. cit.*) with 50 mg ampicillin was used for selection (without glycerol-3-phosphate). Retransformation of BB26-36 was performed by the method of D. Hanahan, J. Mol. Biol. 166 (1983) pages 557-580, with subsequent plating on the selective medium mentioned.

Delection clones for determining the sequence of the DNA fragments of positive cDNA clones were produced by means of exonuclease III (Strategene) and were sequenced according to the method of Sanger et al., Proc. Nat. Acad. Sci. 74 (1977) pages 5463-5467. Some of the DNA sequencing was performed radioactively with the help of the T7 Sequencing® Kit or with a Pharmacia Automated Laser Fluorescent A.L.F.® DNA sequencer. The sequences were analyzed with the help of computer software from the University of Wisconsin Genetics Computer Group (J. Devereux et al., Nucl. Acids Res. 12 (1984) pages 387-394).

Furthermore, cDNA clones were isolated by screening a cDNA library from **Cuphea lanceolata** with ClGPDH20 cDNA as a probe. For this, a cDNA library from **Cuphea lanceolata** (wild type) was produced according to the manufacturer's instructions with the ZAP® cDNA Synthesis Kit. Messenger RNA from isolated, immature embryos about two to three weeks old was the raw material for synthesis of the cDNAs. The cDNA library obtained contained 9.6×10^5 recombinant phages with approx. 50% clones with more than 500 bp insertions. The cDNA library was examined with CLGPDH20 as a probe, and 18 cDNAs were isolated and partially or completely sequenced in the usual manner. Of these cDNAs, 12 were class A, and 6 cDNAs were in class B.

The enzyme measurements were performed with the fusion protein according to the method of Santora et al., Arch. Biochem. Biophys. 196 (1979) pages 403-411.

Example 2

Production of genomic clones of glycerol-3-phosphate dehydrogenase from Cuphea lanceolata

Genomic DNA from young **Cuphea lanceolata** leaves were isolated for this example (S.L. Della Porta et al., Plant. Mol. Biol. Rep. 1, (1983) pages 19-21). The DNA was then partially cleaved with the restriction enzyme Sau3A, whereupon DNA fragments of 11,000 to 19,000 base pairs were cloned in vector lFIXII (Stratagene) that was cleaved with XhoI after the respective interfaces were partially filled with two nucleotides in any given case. The genomic DNA library that was not reproduced amounted to 5.4 times the genome of **Cuphea lanceolata**. Thirty-one genomic clones were then isolated from this library with ClGPDH20-cDNA as a probe.

The three genomic clones ClGPDHg3 (15.9 kb DNA insertion), ClGPDHg5 (17.7 kb DNA insertion) and ClGPDHg9 (15.6 kb DNA insertion) were characterized in greater detail. Suitable subclones were produced in the usual manner and their insertions were sequenced with the ExoIII/Mung bean kit and also with oligonucleotide primers in order to bridge any gaps.

If any of the procedures customary in molecular biology have not have been described adequately here, such procedures were performed by standard methods as described in Sambrook et al., A Laboratory Manual, second edition (1989).

SEQUENCE LIST

(1) GENERAL INFORMATION:

(i) APPLICANT:

(A) NAME: Max Planck Society for Promotion of the Sciences E.V.

(B) STREET: Bunsenstrasse 10

(C) CITY: Goettingen

(E) COUNTRY: Germany

(F) ZIP: 37073

(ii) TITLE OF INVENTION:

Glycerol-3-phosphate dehydrogenase (GPDH)

(iii) NUMBER OF SEQUENCES: 8

(iv) COMPUTER-READABLE FORM:

(A) MEDIUM TYPE: 3.5 inch HD diskette (1.44 MB)/ ASCII Format

(B) COMPUTER: IBM compatible PC

(C) OPERATING SYSTEM: PC-DOS/MS-DOS

(D) SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPA)

(2) INFORMATION FOR ID SEQ NO:1

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 1354 base pairs

(B) TYPE: Nucleic acid

(C) STRANDEDNESS: Double strand

(D) TOPOLOGY: Linear

(ii) MOLECULE TYPE: cDNA to mRNA

(iii) HYPOTHETICAL: No

(iv) ANTI-SENSE: No

(vi) ORIGINAL SOURCE:

(A) ORGANISM: Cuphea lanceolata

(vii) IMMEDIATE SOURCE:

(A) LIBRARY: ZAP cDNA library

(B) CLONE: ClGPDH20

(ix) FEATURE:

(A) NAME/KEY: cDNA

(B) LOCATION: 15 to 1345

(ix) FEATURE:

(A) NAME/KEY: Fusion with lacZ

(B) LOCATION: 1 to 14

(ix) FEATURE:

(A) NAME/KEY: Start codon

(B) LOCATION: 17 to 19

(ix) FEATURE:

(A) NAME/KEY: Stop codon

(B) LOCATION: 1133 to 1135

(ix) FEATURE:

(A) NAME/KEY: PolyA signal

(B) LOCATION: 1329 to 1334

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:1

```
GAATTCGGCA CGAGCA ATG GCT CCC TCT GAG CTC AAC TGC ACC CAC CAG      49
                  Met Ala Pro Ser Glu Leu Asn Cys Thr His Gln
                    1               5                  10

AAC CAG CAT TCA AGC GGT TAC GAC GGA CCC AGA TCG AGG GTC ACC GTT    97
Asn Gln His Ser Ser Gly Tyr Asp Gly Pro Arg Ser Arg Val Thr Val
          15                  20                  25
```

```
GTC GGT AGT GGA AAC TGG GGT AGT GTT GCT GCC AAG CTC ATT GCT ACC   145
Val Gly Ser Gly Asn Trp Gly Ser Val Ala Ala Lys Leu Ile Ala Thr
         30                  35                  40

AAT ACC CTC AAG CTT CCA TCT TTT CAT GAT GAA GTG AGA ATG TGG GTA   193
Asn Thr Leu Lys Leu Pro Ser Phe His Asp Glu Val Arg Met Trp Val
     45                  50                  55

TTT GAG GAG ACG CTA CCG AGC GGC GAG AAG CTT ACT GAT GTC ATC AAC   241
Phe Glu Glu Thr Leu Pro Ser Gly Glu Lys Leu Thr Asp Val Ile Asn
 60                  65                  70                  75

CAG ACC AAT GAA AAT GTT AAG TAT CTC CCC GGA ATT AAG CTC GGT AGG   289
Gln Thr Asn Glu Asn Val Lys Tyr Leu Pro Gly Ile Lys Leu Gly Arg
                 80                  85                  90

AAT GTT GTT GCA GAT CCA GAC CTC GAA AAC GCA GTT AAG GAT GCA AAT   337
Asu Val Val Ala Asp Pro Asp Leu Glu Asn Ala Val Lys Asp Ala Asn
             95                  100                 105

ATG CTC GTG TTT GTG ACA CCG CAT CAG TTC ATG GAG GGC ATC TGC AAA   385
Met Leu Val Phe Val Thr Pro His Gln Phe Met Glu Gly Ile Cys Lys
         110                 115                 120

AGA CTC GAA GGG AAA ATA CAA GAA GGA GCA CAG GCT CTC TCC CTT ATA   433
Arg Leu Glu Gly Lys Ile Gln Glu Gly Ala Gln Ala Leu Ser Leu Ile
     125                 130                 135

AAG GGC ATG GAG GTC AAA ATG GAG GGG CCT TGC ATG ATC TCG AGC TTA   481
Lys Gly Met Glu Val Lys Met Glu Gly Pro Cys Met Ile Ser Ser Leu
140                 145                 150                 155

ATC TCT GAT CTT CTC GGG ATT AAC TGC TGT GTC CTA ATG GGG GCA AAC   529
Ile Ser Asp Leu Leu Gly Ile Asn Cys Cys Val Leu Met Gly Ala Asn
                160                 165                 170

ATC GCT AAT GAG ATT GCT GTT GAG AAA TTC AGT GAA GCG ACA GTC GGG   577
Ile Ala Asn Glu Ile Ala Val Glu Lys Phe Ser Glu Ala Thr Val Gly
            175                 180                 185

TTC AGA GAA AAT AGA GAT ATT GCA GAG AAA TGG GTT CAG CTC TTT AGC   625
Phe Arg Glu Asn Arg Asp Ile Ala Glu Lys Trp Val Gln Leu Phe Ser
        190                 195                 200

ACT CCG TAC TTC ATG GTC TCA GCT GTT GAA GAT GTT GAA GGA GTA GAA   673
Thr Pro Tyr Phe Met Val Ser Ala Val Glu Asp Val Glu Gly Val Glu
    205                 210                 215

CTT TGT GGA ACA CTG AAG AAT ATC GTG GCC ATA GCA GCC GGT TTT GTG   721
Leu Cys Gly Thr Leu Lys Asn Ile Val Ala Ile Ala Ala Gly Phe Val
220                 225                 230                 235

GAT GGA TTG GAG ATG GGA AAC AAC ACA AAA GCA GCA ATT ATG AGG ATC   769
Asp Gly Leu Glu Met Gly Asn Asn Thr Lys Ala Ala Ile Met Arg Ile
```

```
                240                 245                 250

GGG TTA CGG GAG ATG AAG GCA TTC TCC AAG CTT TTG TTT CCA TCT GTT    817
Gly Leu Arg Glu Met Lys Ala Phe Ser Lys Leu Leu Phe Pro Ser Val
            255                 260                 265

AAG GAC ACT ACT TTC TTC GAG AGC TGT GGA GTC GCT GAC CTC ATC ACA    865
Lys Asp Thr Thr Phe Phe Glu Ser Cys Gly Val Ala Asp Leu Ile Thr
        270                 275                 280

ACT TGT TTG GGC GGG AGA AAC AGA AAA GTT GCT GAG GCT TTT GCA AAG    913
Thr Cys Leu Gly Gly Arg Asn Arg Lys Val Ala Glu Ala Phe Ala Lys
    285                 290                 295

AAT GGC GGG AAA AGG TCA TTC GAT GAT CTC GAA GCA GAG ATG CTC CGG    961
Asn Gly Gly Lys Arg Ser Phe Asp Asp Leu Glu Ala Glu Met Leu Arg
300                 305                 310                 315

GGG CAA AAA TTA CAG GGT GTC TCA ACA GCA AAG GAG GTC TAT GAA GTC   1009
Gly Gln Lys Leu Gln Gly Val Ser Thr Ala Lys Glu Val Tyr Glu Val
                320                 325                 330

TTG GGG CAC CGA GGC TGG CTC GAG CTG TTC CCG CTC TTC TCA ACC GTG   1057
Leu Gly His Arg Gly Trp Leu Glu Leu Phe Pro Leu Phe Ser Thr Val
            335                 340                 345

CAC GAG ATA TCC ACT GGC CGT CTG CCT CCT TCA GCC ATC GTC GAA TAC   1105
His Glu Ile Ser Thr Gly Arg Leu Pro Pro Ser Ala Ile Val Glu Tyr
        350                 355                 360

AGC GAA CAA AAA ACC ATC TTC TCT TGG TAGAGCAAGA GGCTGCCCTT         1152
Ser Glu Gln Lys Thr Ile Phe Ser Trp
    365                 370

GAAAGACTAA GAGCCACCCT GCCCTGTTTA AAGGGCTAAA AGTTTAATAT TTCTCTGCAG 1212

CCTAAACAGT CGGAAACATT GAAAATCTAG GATGTATAAG AAAAAAAAAA GAAGGTTTGA 1272

AGGAAGTATG GATGGGCATG AATGTATTTA TTTTCGGTAT ACTCTTTTTC TGCAAAAATA 1332

ATTTCTTCAG AAAGGGGGGC CC                                          1354
```

(2) INFORMATION FOR ID SEQ NO:2

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 1464 base pairs

(B) TYPE: Nucleic acid

(C) STRANDEDNESS: Double stranded

(D) TOPOLOGY: Linear

(ii) MOLECULE TYPE: cDNA to mRNA

(iii) HYPOTHETICAL: No

(iv) ANTI-SENSE: No

(vi) ORIGINAL SOURCE:

(A) ORGANISM: **Cuphea lanceolata**

(vii) IMMEDIATE SOURCE:

(A) LIBRARY: ZAP cDNA library

(B) CLONE: ClGPDH109

(ix) FEATURE:

(A) NAME/KEY: cDNA

(B) LOCATION: 15 to 1454

(ix) FEATURE:

(A) NAME/KEY: CDS [coding sequence]

(B) LOCATION: 15 to 1187

(ix) FEATURE:

(A) NAME/KEY: Fusion with lacZ

(B) LOCATION: 1 to 14

(ix) FEATURE:

(A) NAME/KEY: Start codon

(B) LOCATION: 45 to 47

(ix) FEATURE:

(A) NAME/KEY: Stop codon

(B) LOCATION: 1188 to 1190

(ix) FEATURE:

(A) NAME/KEY: PolyA signal

(B) LOCATION: 1414 to 1419

(ix) FEATURE:

(A) NAME/KEY: PolyA region

(B) LOCATION: 1446 to 1454

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:2

```
GAATTCGGCA CGAGCTTCCT CTGTTCTTCC TCTCTGCCTC TGCA ATG GCG CCT GCC     56
                                                 Met Ala Pro Ala
                                                   1

TTC GAA CCC CAT CAG CTG GCT CCC TCT GAG CTT AAC TCT GCC CAC CAG     104
Phe Glu Pro His Gln Leu Ala Pro Ser Glu Leu Asn Ser Ala His Gln
  5                  10                  15                  20

AAC CCA CAT TCA GGC GGA TAT GAC GGA CCC AGA TCG AGG GTC ACT GTC     152
Asn Pro His Ser Gly Gly Tyr Asp Gly Pro Arg Ser Arg Val Thr Val
                 25                  30                  35

GTC GGC AGC GGC AAC TGG GGC AGC GTC GCT GCC AAG CTC ATT GCT TCC     200
Val Gly Ser Gly Asn Trp Gly Ser Val Ala Ala Lys Leu Ile Ala Ser
             40                  45                  50

AAC ACC CTC AAG CTC CCA TCT TTC CAT GAT GAA GTG AGG ATG TGG GTA     248
Asu Thr Leu Lys Leu Pro Ser Phe His Asp Glu Val Arg Met Trp Val
         55                  60                  65

TTT GAG GAG ACT CTA CCG GGC GGC GAG AAG CTC ACT GAT ATC ATC AAC     296
Phe Glu Glu Thr Leu Pro Gly Gly Glu Lys Leu Thr Asp Ile Ile Asn
     70                  75                  80

CAG ACC AAT GAA AAT GTT AAA TAT CTT CCC GGA ATT AAG CTC GGT GGG     344
Glu Thr Asn Glu Asn Val Lys Tyr Leu Pro Gly Ile Lys Leu Gly Gly
85                  90                  95                  100

AAT GTT GTT GCT GAT CCA GAC CTC GAA AAT GCA GTT AAG GAT GCA AAT     392
Asn Val Val Ala Asp Pro Asp Leu Glu Asn Ala Val Lys Asp Ala Asn
                105                 110                 115

ATG CTC GTG TTT GTC ACA CCG CAT CAG TTC ATG GAG GGC ATC TGC AAA     440
Met Leu Val Phe Val Thr Pro His Gln Phe Met Glu Gly Ile Cys Lys
            120                 125                 130

AGA CTT GTC GGG AAG ATA CAG GAA GGA GCG CAG GCT CTC TCC CTT ATA     488
Arg Leu Val Gly Lys Ile Gln Glu Gly Ala Gln Ala Leu Ser Leu Ile
        135                 140                 145

AAA GGC ATG GAG GTC AAG ATG GAG GGG CCT TGC ATG ATC TCG AGC CTA     536
Lys Gly Met Glu Val Lys Met Glu Gly Pro Cys Met Ile Ser Ser Leu
    150                 155                 160

ATC TCA GAT CTT CTC GGG ATC AAC TGC TGT GTC CTT AAT GGG GCA AAC     584
Ile Ser Asp Leu Leu Gly Ile Asu Cys Cys Val Leu Asn Gly Ala Asn
```

```
165                 170                 175                 180

ATC GCT AAT GAG ATT GCT GTT GAG AAA TTC AGT GAA GCG ACT GTC GGG   632
Ile Ala Asn Glu Ile Ala Val Glu Lys Phe Ser Glu Ala Thr Val Gly
                185                 190                 195

TTC AGA GAA AAT AGA GAT ATT GCG GAA AAA TGG GTT CAG CTC TTT AGC   680
Phe Arg Glu Asn Arg Asp Ile Ala Glu Lys Trp Val Gln Leu Phe Ser
            200                 205                 210

ACT CCA TAC TTC ATG GTC TCA GCT GTT GAA GAT GTT GAA GGA GTA GAG   728
Thr Pro Tyr Phe Met Val Ser Ala Val Glu Asp Val Glu Gly Val Glu
        215                 220                 225

CTT TGT GGA ACA CTG AAG AAT ATT GTG GCC ATA GCA GCG GGT TTT GTT   776
Leu Cys Gly Thr Leu Lys Asn Ile Val Ala Ile Ala Ala Gly Phe Val
    230                 235                 240

GAT GGA TTG GAG ATG GGA AAC AAC ACA AAA GCG GCA ATT ATG AGG ATC   824
Asp Gly Leu Glu Met Gly Asn Asn Thr Lys Ala Ala Ile Met Arg Ile
245                 250                 255                 260

GGG CTG CGG GAG ATG AAA GCG TTC TCC AAG CTT TTG TTT CCA TCT GTT   872
Gly Leu Arg Glu Met Lys Ala Phe Ser Lys Leu Leu Phe Pro Ser Val
                265                 270                 275

AAG GAC ACT ACT TTT TTC GAG AGC TGC GGA GTC GCT GAT CTC ATC ACA   920
Lys Asp Thr Thr Phe Phe Glu Ser Cys Gly Val Ala Asp Leu Ile Thr
            280                 285                 290

ACT TGT TTG GGC GGA AGA AAC AGA AAA GTC GCT GAG GCT TTT GCA AAG   968
Thr Cys Leu Gly Gly Arg Asn Arg Lys Val Ala Glu Ala Phe Ala Lys
        295                 300                 305

AAT GGC GGA AAC AGG TCA TTT GAT GAT CTC GAA GCA GAG ATG CTC CGG  1016
Asn Gly Gly Asn Arg Ser Phe Asp Asp Leu Glu Ala Glu Met Leu Arg
    310                 315                 320

GGG CAA AAA TTA CAG GGT GTC TCG ACA GCG AAA GAG GTC TAC GAG GTC  1064
Gly Gln Lys Leu Gln Gly Val Ser Thr Ala Lys Glu Val Tyr Glu Val
325                 330                 335                 340

CTG AGG CAC CGA GGC TGG CTC GAG TTG TTC CCG CTC TTC TCA ACC GTG  1112
Leu Arg His Arg Gly Trp Leu Glu Leu Phe Pro Leu Phe Ser Thr Val
                345                 350                 355

CAT GAG ATC TCC AGT GGC CGT CTG CCT CCT TCA GCC ATT GTT GAA TAC  1160
His Glu Ile Ser Ser Gly Arg Leu Pro Pro Ser Ala Ile Val Glu Tyr
            360                 365                 370

AGC GAA CAA AAG CCT ACC TTC TCT TGG TAGAGAAAGA AACCAGGAAG        1207
Ser Glu Gln Lys Pro Thr Phe Ser Trp
        375                 380
```

```
AACGGCGAGC CACTGTCCCC CGTTTAAAGG TTTACTATTT CTCTCTGCAC TTTGCAGCCT 1267
GAAGAGTCGG AAACATAGAA AATCTAGGAA GTTTCAGAAA AAGGAAGGTT TTGAGGATGT 1327
ATGGATGATA TATATACTAG GTGGGTATGA AGAGGAAGTT ATTACTATGA TGTTGGTATG 1387
TGGTAATGGC TAAGTACATG AGATCAAATA AATAGACAGA CCTTGGTTTC TTCTTTCTAA 1447
AAAAAAAGGG GGGGCCC                                                 1464
```

(2) INFORMATION FOR ID SEQ NO:3

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 1490 base pairs

(B) TYPE: Nucleic acid

(C) STRANDEDNESS: Double

(D) TOPOLOGY: Linear

(ii) MOLECULE TYPE: cDNA to mRNA

(iii) HYPOTHETICAL: No

(iv) ANTI-SENSE: No

(vi) ORIGINAL SOURCE:

(A) ORGANISM: **Cuphea lanceolata**

(vii) IMMEDIATE SOURCE:

(A) LIBRARY: ZAP cDNA library

(B) CLONE:ClGPDH132

(ix) FEATURE:

(A) NAME/KEY: CDS

(B) LOCATION: 15 to 1115

(ix) FEATURE:

(A) NAME/KEY: Fusion with lacZ

(B) LOCATION: 1 to 14

(ix) FEATURE:

(A) NAME/KEY: Stop codon

(B) LOCATION: 1116 to 1118

(ix) FEATURE:

(A) NAME/KEY: PolyA signal

(B) LOCATION: 1343 to 1348

(ix) FEATURE:

(A) NAME/KEY: PolyA region

(B) LOCATION: 1465 to 1484

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:3

```
GAATTCGGCA CGAG CTT AAC TCT GCC CAC CAG AAC CCA CAT TCC AGC GGA      50
               Leu Asn Ser Ala His Gin Asn Pro His Ser Ser Gly
                 1               5                  10

TAT GAA GGA CCC AGA TCG AGG GTC ACC GTC GTT GGC AGC GGC AAC TGG      98
Tyr Glu Gly Pro Arg Ser Arg Val Thr Val Val Gly Ser Gly Asn Trp
         15                  20                  25

GGC AGC GTC GCT GCC AAG CTC ATT GCT TCC AAC ACC CTC AAG CTC CCA     146
Gly Ser Val Ala Ala Lys Leu Ile Ala Ser Asn Thr Leu Lys Leu Pro
     30                  35                  40

TCT TTC CAT GAT GAA GTG AGG ATG TGG GTA TTT GAG GAG ACT CTA CCG     194
Ser Phe His Asp Glu Val Arg Met Trp Val Phe Glu Glu Thr Leu Pro
 45                  50                  55                  60

GGC GGC GAG AAG CTC ACT GAT ATC ATC AAC CAG ACC AAT GAA AAT GTT     242
Gly Gly Glu Lys Leu Thr Asp Ile Ile Asn Gln Thr Asn Glu Asn Val
                 65                  70                  75

AAA TAT CTT CCC GGA ATT AAG CTC GGT AGG AAT GTT GTT GCA GAT CCA     290
Lys Tyr Leu Pro Gly Ile Lys Leu Gly Arg Asn Val Val Ala Asp Pro
             80                  85                  90

GAC CTC GAA AAC GCA GTT AAG GAT GCA AAT ATG CTC GTT TTC GTC ACA     338
Asp Leu Glu Asn Ala Val Lys Asp Ala Asn Met Lou Val Phe Val Thr
         95                  100                 105

CCG CAT CAG TTC GTG GAG GGC ATC TGC AAA AGA CTT GTA GGG AAG ATA     386
Pro His Gin Phe Val Glu Gly Ile Cys Lys Arg Leu Val Gly Lys Ile
     110                 115                 120

CAG GAA GGA GCG CAG GCT CTC TCT CTT ATA AAA GGC ATG GAG GTC AAA     434
Gin Glu Gly Ala Gin Ala Leu Ser Leu Ile Lys Gly Met Glu Val Lys
125                 130                 135                 140

ATG GAG GGG CCT TGC ATG ATC TCG AGC CTA ATC TCA GAT CTT CTC GGG     482
Met Glu Gly Pro Cys Met Ile Ser Ser Leu Ile Ser Asp Leu Leu Gly
                145                 150                 155
```

```
ATC AAT TGC TGT GTC CTT AAT GGG GCG AAC ATC GCT AAT GAG ATT GCT     530
Ile Asn Cys Cys Val Leu Asn Gly Ala Asn Ile Ala Asn Glu Ile Ala
            160                 165                 170

GTT GAG AAA TTC AGT GAA GCG ACT GTC GGG TTC AGA GAA AAT AGA GAT     578
Val Glu Lys Phe Ser Glu Ala Thr Val Gly Phe Arg Glu Asn Arg Asp
        175                 180                 185

ATT GCG GAA AAA TGG GTT CAG CTC TTT AGC ACT CCA TAC TTC ATG GTC     626
Ile Ala Glu Lys Trp Val Gln Leu Phe Ser Thr Pro Tyr Phe Met Val
    190                 195                 200

TCA GCT GTT GAA GAT GTT GAA GGA GTA GAG CTT TGT GGA ACA CTG AAG     674
Ser Ala Val Glu Asp Val Glu Gly Val Glu Leu Cys Gly Thr Leu Lys
205                 210                 215                 220

AAT ATT GTG GCC ATA GCA GCG GGT TTT GTG GAT GGA CTG GAG ATG GGA     722
Asu Ile Val Ala Ile Ala Ala Gly Phe Val Asp Gly Leu Glu Met Gly
                225                 230                 235

AAC AAC ACA AAA GCA GCA ATT ATG AGG ATC GGG CTG CGG GAG ATG AAA     770
Asn Asn Thr Lys Ala Ala Ile Met Arg Ile Gly Leu Arg Glu Met Lys
            240                 245                 250

GCG TTC TCC AAG CTT TTG TTT CCA TCT GTT AAG GAC ACT ACT TTT TTC     818
Ala Phe Ser Lys Leu Leu Phe Pro Ser Val Lys Asp Thr Thr Phe Phe
        255                 260                 265

GAG AGC TGC GGA GTC GCT GAT CTC ATC ACA ACT TGT TTG GGC GGA AGA     866
Glu Ser Cys Gly Val Ala Asp Leu Ile Thr Thr Cys Leu Gly Gly Arg
    270                 275                 280

AAC AGA AAA GTC GCT GAG GCT TTT GCA AAG AAT GGC GGT AAC AGG TCA     914
Asn Arg Lys Val Ala Glu Ala Phe Ala Lys Asn Gly Gly Asn Arg Ser
285                 290                 295                 300

TTC GAT GAT CTC GAA GCA GAG ATG CTC CGG GGG CAA AAA TTA CAG GGT     962
Phe Asp Asp Leu Glu Ala Glu Met Leu Arg Gly Gln Lys Leu Gln Gly
                305                 310                 315

GTC TCG ACA GCG AAA GAG GTC TAC GAG GTC CTG AGG CAC CGA GGT TGG    1010
Val Ser Thr Ala Lys Glu Val Tyr Glu Val Leu Arg His Arg Gly Trp
            320                 325                 330

CTC GAG TTG TTC CCG CTC TTC TCA ACC GTG CAT GAG ATC TCC ACT GGC    1058
Leu Glu Leu Phe Pro Leu Phe Ser Thr Val His Glu Ile Ser Thr Gly
        335                 340                 345

CGT CTG CCT CCT TCA GCC ATT GTT GAA TAC AGC GAA CAA AAG CCC ACC    1106
Arg Leu Pro Pro Ser Ala Ile Val Glu Tyr Ser Glu Gln Lys Pro Thr
    350                 355                 360

TTC TCT TGG TAGAGAAAGA AGCAACCAGG AAGAACGGCG AGCCACTCTG            1155
Phe Ser Trp
```

365

```
CCTCGTTTAA AGGGTTACTA TTTCTCTACA CTCTGCAGCC TGAAGAGTCG GAAACATCGA 1215
AAATCTAGGA AGTCTCAGAA AAATGAAGGT TTGGAGGATG TATGGATGAT ATATATACTA 1275
GGTGGGTATG AAGAGGAAGT TATTACTATG ATGTTGGTAT GTGGTAATGG CTAAGTACAT 1335
GAGATCAAAT AAATAGACAG ACCTTGGTTT CTTCTATCTC GATTCGGTCT CGTCGAGTTT 1395
GGCGAAACTC AACTGAACTT CCTGAGTACC CTGCTACCTA TTACATGTAA TGTTCCTATT 1455
TATATGCTTA AAAAAAAAAA AAAAAAAAAC TCGAG                             1490
```

(2) INFORMATION FOR ID SEQ NO:4

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 1390 base pairs

(B) TYPE: Nucleic acid

(C) STRANDEDNESS: Double strand

(D) TOPOLOGY: Linear

(ii) MOLECULE TYPE: cDNA to mRNA

(iii) HYPOTHETICAL: No

(iv) ANTI-SENSE: No

(vi) ORIGINAL SOURCE:

(A) ORGANISM: **Cuphea lanceolata**

(vii) IMMEDIATE SOURCE:

(A) LIBRARY: ZAP cDNA library

(B) CLONE:ClGPDH30

(ix) FEATURE:

(A) NAME/KEY: cDNA

(B) LOCATION: 15 to 1384

(ix) FEATURE:

(A) NAME/KEY: CDS

(B) LOCATION: 34 to 1149

(ix) FEATURE:

(A) NAME/KEY: Fusion with lacZ

(B) LOCATION: 1 to 14

(ix) FEATURE:

(A) NAME/KEY: Start codon

(B) LOCATION: 34 to 36

(ix) FEATURE:

(A) NAME/KEY: Stop codon

(B) LOCATION: 1150 to 1152

(ix) FEATURE:

(A) NAME/KEY: PolyA signal

(B) LOCATION: 1349 to 1354

(ix) FEATURE:

(A) NAME/KEY: PolyA region

(B) LOCATION: 1366 to 1384

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:4

```
GAATTCGGCA CGAGTTTCTT CTCAGCCTCT GCA ATG GCT CCC TCT GAG CTC AAC      54
                                      Met Ala Pro Ser Glu Leu Asn
                                        1               5

TGC ACC CAC CAG AAC CCA CAT TCA AGC GGT TAC GAC GGA CCC AGA TCG     102
Cys Thr His Gln Asn Pro His Ser Ser Gly Tyr Asp Gly Pro Arg Ser
         10                  15                  20

AGG GTC ACC GTT GTC GGT AGT GGA AAC TGG GGC AGT GTC GCT GCC AAG     150
Arg Val Thr Val Val Gly Ser Gly Asn Trp Gly Ser Val Ala Ala Lys
     25                  30                  35

CTC ATT GCT TCC AAT ACC CTC AAG CTT CCA TCT TTT CAT GAT GAA
Leu Ile Ala Ser Asn Thr Leu Lys Leu Pro Ser Phe His Asp Glu
 40                  45                  50

AGA ATG TGG GTA TTT GAG GAG ACT CTA CCG AGC GGC GAG AAG CTT ACT
Arg Met Trp Val Phe Glu Glu Thr Leu Pro Ser Gly Glu Lys Leu Thr
                 60                  65                  70

GAT GTC ATC AAC CAG ACC AAT GAA AAT GTT AAG TAT CTC CCC GGA ATT     294
Asp Val Ile Asn Gln Thr Asn Glu Asn Val Lys Tyr Leu Pro Gly Ile
             75                  80                  85

AAG CTC GGT AGG AAT GTT GTT GCA GAT CCA GAC CTC GAA AAC GCA GTT     342
Lys Leu Gly Arg Asn Val Val Ala Asp Pro Asp Leu Glu Asn Ala Val
```

```
          90                  95                 100

AAG GAT GCA AAT ATG CTC GTG TTT GTG ACA CCG CAT CAG TTC ATG GAG      390
Lys Asp Ala Asn Met Leu Val Phe Val Thr Pro His Gln Phe Met Glu
    105                 110                 115

GGC ATC TGC AAA AGA CTC GTA GGG AAA ATA CAG GAA GGA GCA CAG GCT      438
Gly Ile Cys Lys Arg Leu Val Gly Lys Ile Gln Glu Gly Ala Gln Ala
120                 125                 130                 135

CTC TCC CTT ATA AAG GGC ATG GAG GTC AAA ATG GAG GGG CCT TGC ATG      486
Leu Ser Leu Ile Lys Gly Met Glu Val Lys Met Glu Gly Pro Cys Met
                140                 145                 150

ATC TCG AGC CTA ATC TCT GAT CTT CTC GGG ATC AAC TGC TGT GTC CTA      534
Ile Ser Ser Leu Ile Ser Asp Leu Leu Gly Ile Asn Cys Cys Val Leu
            155                 160                 165

ATG GGG GCA AAC ATC GCT AAT GAG ATT GCT GTT GAG AAA TTC AGT GAA      582
Met Gly Ala Asn Ile Ala Asn Glu Ile Ala Val Glu Lys Phe Ser Glu
        170                 175                 150

GCG ACA GTC GGG TTC AGA GAA AAT ACA GAT ATT GCG GAG AAA TGG GTT      630
Ala Thr Val Gly Phe Arg Glu Asn Thr Asp Ile Ala Glu Lys Trp Val
    185                 190                 195

CAG CTC TTT AGC ACT CCG TAC TTC ATG GTC TCA GCT GTT GAA GAT GTT      678
Gln Leu Phe Ser Thr Pro Tyr Phe Met Val Ser Ala Val Glu Asp Val
200                 205                 210                 215

GAA GGA GTA GAA CTT TGT GGA ACA CTG AAG AAT ATC GTG GCC ATA GCA      726
Glu Gly Val Glu Leu Cys Gly Thr Leu Lys Asn Ile Val Ala Ile Ala
                220                 225                 230

GCC GGT TTT GTG GAT GGA TTG GAG ATG GGA AAC AAC ACA AAA GCA GCA      774
Ala Gly Phe Val Asp Gly Leu Glu Met Gly Asn Asn Thr Lys Ala Ala
            235                 240                 245

ATT ATG AGG ATC GGG TTA CGG GAG ATG AAG GCA TTC TCC AAG CTT TTG      822
Ile Met Arg Ile Gly Leu Arg Glu Met Lys Ala Phe Ser Lys Leu Leu
        250                 255                 260

TTT CCA TCT GTT AAG GAC ACT ACT TTC TTC GAG AGC TGT GGA GTT GCT      870
Phe Pro Ser Val Lys Asp Thr Thr Phe Phe Glu Ser Cys Gly Val Ala
    265                 270                 275

GAC CTC ATC ACA ACT TGT TTG GGC GGG AGA AAC AGA AAA GTT GCT GAG      918
Asp Leu Ile Thr Thr Cys Leu Gly Gly Arg Asn Arg Lys Val Ala Glu
280                 285                 290                 295

GCT TTT GCA AAG KAT GGC GGG GAA AGG TCA TTC GAT GAT CTC GAA GCA      966
Ala Phe Ala Lys Asn Gly Gly Glu Arg Ser Phe Asp Asp Leu Glu Ala
                300                 305                 310
```

```
GAG CTG CTC CGG GGG CAA AAA TTA CAG GGT GTC TCA ACA GCA AAG GAG     1014
Glu Leu Leu Arg Gly Gln Lys Lou Gln Gly Val Ser Thr Ala Lys Glu
            315                 320                 325

GTC TAT GAA GTC TTG GGG CAC CGA GGC TGG CTC GAG CTG TTC CCG CTC     1062
Val Tyr Glu Val Leu Gly His Arg Gly Trp Leu Glu Leu Phe Pro Leu
        330                 335                 340

TTC TCA ACC GTG CAC GAG ATC TCC ACT GGC CGT CTG CAT CCT TCA GCC     1110
Phe Ser Thr Val His Glu Ile Ser Thr Gly Arg Leu His Pro Ser Ala
    345                 350                 355

ATC GTC GAA TAC AGC GAA CAA AAA ACC ATC TTC TCT TGG TAGAGCAAGA      1159
Ile Val Glu Tyr Ser Glu Gln Lys Thr Ile Phe Ser Trp
360                 365                 370

GGCTGCCCTT GAAAGACTAA GAGCCACCCT GCCCTGTTTA AAGGGCTAAA AGTTTAATAT   1219
TTCTCTGCAG CCTAAACAGT TGGAAACATT GAAAATCTAG GATGTATCAG AAAAAAGAAG   1279
GTTTGGAGGA AGTATGGATG ATATAGAGGA CATGAATGTA TTCATTTTCG GTATACTCTT   1339
TTTCTGCAAA ATAATTCTTC AGATGTAAAA AAAAAAAAAA AAAAACTCGA G            1390
```

(2) INFORMATION FOR ID SEQ NO:5

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 4434 base pairs

(B) TYPE: Nucleic acid

(C) STRANDEDNESS: Double strand

(D) TOPOLOGY: Linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: No

(iv) ANTI-SENSE: No

(vi) ORIGINAL SOURCE:

(A) ORGANISM: **Cuphea lanceolata**

(vii) IMMEDIATE SOURCE:

(A) LIBRARY: Genomic lambda FIX II

(B) CLONE: ClGPDHg5

(ix) FEATURE:

(A) NAME/KEY: TATA signal

(B) LOCATION: 1332 to 1336

(ix) FEATURE:

(A) NAME/KEY: Start codon

(B) LOCATION: 1394 to 1396

(ix) FEATURE:

(A) NAME/KEY: CDS

(B) LOCATION: Join (1394 to 1550, 2066 to 2142, 2241 to 2313, 2405 to 2622, 2719 to 2826, 2961 to 3024, 3233 to 3260, 3342 to 3462, 3541 to 3595, 3692 to 3740, 3580 to 4005)

(ix) FEATURE:

(A) NAME/KEY: Stop codon

(B) LOCATION: 4006 to 4008

(ix) FEATURE:

(A) NAME/KEY: PolyA signal

(B) LOCATION: 4205 to 4210

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:5

```
GGATCCTTAG AAGACAAGCG CGGGGCGGGC ATGGGTCTCG TGATACCCGC CCCATTTTGC     60
CCCATTCCAT CCCTATATGG TAAGCAGATC  TCACTGAAAA GTCACCGTTT CTGGATGGTT   120
TCCAGATGAT TTTGTCCCTC CCTCTAGCTG CATTAGGTGA TGGGATTGAG GCTATTCTAA   180
GAACCAGCTC GTGTGGAAGG TAGGCGGAGA TTAGCTCCCA GTTCCATCCT CCTGTATTTG   240
AAGCGAAGAA AGAAACTGGG TTGTCTAGCA TGTTTTGTGG GACAGGTTTG GTCGTCTTTT   300
CTGATAGGCT CTGATTCAAT AGAAGCCAAT TATCTCTCCA AAAGGAAACC TTATTACCAC   360
TTCCAATCGA CCACCCTATG TACTTGCTGA TCTTCGGCCA GGTATCGCAT AAAGCATTCC   420
ATAACGCTGA TGCTGTCGTC TTTTTTGTGA ATGTTGGCAA GAGTGTGTCT GGCATGGCAT   480
ATTTGTGACT GAGCACCCGC ACCCAAAGGC TCTGAGGTTG TGATGCCATA TCCCAACATA   540
CCTTCGATAG AAAGGCTTCA TTCATCTTCC GTAGCTTACG AATGCCAAGA CCACCCCATG   600
GTGCTGGACT AGTGACCGTG GACCAATTGA CCAAATGCAC CTTCCTTTGC TCCATTGAAT   660
GGCCCCAAAT GAAGTTGCCG CAATGTCTTT CGATTTCATC AAGTGTTCCA TGAGGAATAC   720
GTGTGGACTG CATGGAGAAG GATGGCAGAG CCGTCAAGAC AGATTTCACC AGCGTCACCC   780
GCCCAGCCAT TGACAGTGTC GATGCCGACC AACCAGCAAG TCTTGCTTTT ACCTCGACAT   840
GTTTTGGATT TTATATACCG GTGGTGATGG TGTTTGAATT AATCATCGTC ATTAATTTAT   900
ACCGTGCAAT ATATATTGCA ACATTCCAAA GTATAATTAA TTTTATATGT CCATTCGTGA   960
CTAATCTTGG AGATAGGGCT TAAATTGTTA TATGATGATA TAGAAGAAGT TGGATAGCAC  1020
ATAAGAACTC TATAAAATGC TTATAGATCA TGGCATCGAA TTCATCCGCT ATATATGAGT  1080
```

```
GAGGAAGAAA CTAATCAAAA CCTCGTATTC ATCGAAACAA CCGTTGAAGT GGTTACACTT   1140
TGAATCCTAA GACATACTTG ACGTCATGAT TCTGTCTCTC TATTCCATTG CATAATAAAT   1200
AAAACAAAGG AAACAAAAGC ATAGAGGAGA TCGCCAGATT CAGCAGTTTC CGCATAGGTT   1260
GCCACGGAGC CTTACATGCC GATGCCTTCC TCTGCCTCCT TCTTCCTCCT GTCTCTCTCT   1320
CTACATCCCC TTATATCCCT TCCTCCTTCC CTCCATCTTC ACCATTCCTC TGTTTTTCTT   1380

CTCAGCCTCT GCA ATG GCT CCC TCT GAG CTC AAC TGC ACC CAC CAG AAC      1429
               Met Ala Pro Ser Glu Leu Asn Cys Thr His Gln Asn
                 1               5                  10

CCA CAT TCA AGC GGT TAC GAC GGA CCC AGA TCG AGG GTC ACC GTT GTC     1477
Pro His Ser Ser Gly Tyr Asp Gly Pro Arg Ser Arg Val Thr Val Val
         15                  20                  25

GGT AGT GGA AAC TGG GGC AGT GTC GCT GCC AAG CTC ATT GCT TCC AAT     1525
Gly Ser Gly Asn Trp Gly Ser Val Ala Ala Lys Leu Ile Ala Ser Asn
     30                  35                  40

ACC CTC AAG CTT CCA TCT TTT CAT G GTTCGTCTCT CCTTTTCTCT             1570
Thr Leu Lys Leu Pro Ser Phe His
45                  50

GAAAAATGAA GCTTTTGCAT GGGATAGTCA CTAGATATGA GCCTCTGTTT GCATGACTGA   1630
AGCGCTTGAG TAACCGAGTT TTTGGAACAA GAGCACAGGT GGTTCCTTTG CATTTTCTTT   1690
GAGGTTCCTT AATCATTCAA TGAAGTAGCG GTTGATCGCT GAGCAATTGA AACTTGTGGA   1750
ATCGAACCTC CAGCCGAGTC TTAGTGTAAT TGCTTTCTGT TTTACTTCAT TCATAGTGGG   1510
AAGGAGTACG AACTGATGAG TGATGTCACA TTTCATTAGT CGGGTTGCGA AAAAACTCAG   1870
TTGACATATT GGTCGAGACT CTGCAGTGTC ATCAGATATG AGTTGGTGTA TTTGTATTGA   1930
CATTTGAATT TGGTATGTGT ATGAATTTTG TTGAATTAAT CACCGCTGTG ATGAAAAGAT   1990
CAGTACTTCT TCGGTCATTT TTCAGGTGGA AGGATGTTGG TTTCTTATAT ATGTAACTTT   2050
ACATGAATTT TTCAG AT GAA GTG AGA ATG TGG GTA TTT GAG GAG ACT CTA      2100
                Asp Glu Val Arg Met Trp Val Phe Glu Glu Thr Leu
                      55                  60

CCG AGC GGC GAG AAG CTT ACT GAT  GTC ATC AAC CAG ACC AAT            2142
Pro Ser Gly Glu Lys Leu Thr Asp Val Ile Asn Gln Thr Asn
65                  70                  75

GTAAGGAAAC ACAGATTAGC AATAGCATGA GCAGTTATTG CTGGTTAAAT ATGCTTGTTA   2202

GCAACTTTCG TGACGGCCTG AGTTTTATAC CTCTGCAG GAA AAT GTT AAG TAT       2255
                                        Glu Asn Val Lys Tyr
                                             80

CTC CCC GGA ATT AAG CTC GGT AGG AAT GTT GTT GCA GAT CCA GAC CTC     2303
Leu Pro Gly Ile Lys Leu Gly Arg Asn Val Val Ala Asp Pro Asp Leu
     85                  90                  95

GAA AAC GCA G GTAGTCCATG TGTTCATTAG AATTCTCTAA TTAATTATTG           2353
Glu Asn Ala
100

TGGTTTATTT CCTTGTCTCT GTGATGATAT TCTGGATGAA ATTTTGTGCA G TT AAG     2409
                                                        Val Lys
```

```
GAT GCA AAT ATG CTC GTG TTT GTG ACA CCG CAT CAG TTC ATG GAG GGC      2457
Asp Ala Asn Met Leu Val Phe Val Thr Pro His Gln Phe Met Glu Gly
105                 110                 115                 120

ATC TGC AAA  AGA CTC GTA GGG AAA ATA CAG GAA GGA GCA CAG GCT CTC    2505
Ile Cys Lys Arg Leu Val Gly Lys Ile Gln Glu Gly Ala Gln Ala Leu
                125                 130                 135

TCC CTT ATA AAG GGC ATG GAG GTC AAA ATG GAG GGG CCT TGC ATG ATC    2553
Ser Leu Ile Lys Gly Met Glu Val Lys Met Glu Gly Pro Cys Met Ile
            140                 145                 150

TCG AGC CTA ATC TCT GAT CTT CTC GGG ATC AAC TGC TGT GTC CTA ATG    2601
Ser Ser Leu Ile Ser Asp Leu Leu Gly Ile Asn Cys Cys Val Leu Met
        155                 160                 165

GGG GCA AAC ATC GCT AAT GAG GTAAACACTT GGCACGATCT GGTTGCAACT       2652
Gly Ala Asn Ile Ala Asn Glu
    170                 175

CCCCCAGGAA ATTGTAGATC CTCATACTGT TAGCATCTTG ATGAGGTTAA ATATCTTATG  2712

TTGTAG ATT GCT GTT GAG AAA TTC AGT GAA GCG ACA GTC GGG TTC AGA      2760
       Ile Ala Val Glu Lys Phe Ser Glu Ala Thr Val Gly Phe Arg
                    180                 185

GAA AAT ACA GAT ATT GCG GAG AAA TGG GTT CAG CTC TTT AGC ACT CCG     2808
Glu Asn Thr Asp Ile Ala Glu Lys Trp Val Gln Leu Phe Ser Thr Pro
190                 195                 200                 205

TAC TTC ATG GTC TCA GCT GTAAGTTGCG ATAAAACCTT ACGTTTTGCT           2856
Tyr Phe Met Val Ser Ala
                210

AATAGAACAC AATGCTAGAA ACTCCCAGAT TTCAATGTTA TGTATTTTGG TGCCCAAAGA  2916

AGCAACTTCT TAACATCTGT GGCTCCTCTT ACTGACAAAA ATAG GTT GAA GAT GTT   2972
                                                Val Glu Asp Val
                                                            215

GAA GGA GTA GAA CTT TGT GGA ACA CTG AAG AAT ATC GTG GCC ATA GCA     3020
Glu Gly Val Glu Leu Cys Gly Thr Leu Lys Asn Ile Val Ala Ile Ala
                220                 225                 230

GCC G GTTCGTGTTT ACGAGATGTA CATTTATGTA TAACAATCTT TCATTTATTC       3074
Ala

ATCGAGATGG GATGCAATAT ATCAATGAGA GGGAAAAGAA AGGGCAAAGG AAAATGCTGT  3134

TGTATTGCAG CTTTAGGCAT TCTTTTCTCT TAATTATTAA CTGTGAAACA CCGAGAAGTA  3194

TTGATGAAGT TAAGAAACGA TGTTACAG GT TTT GTG GAT GGA TTG GAG ATG       3245
```

```
                                    Gly Phe Val Asp Gly Leu Glu Met
                                    235                 240

GGA AAC AAC ACA AAA GTAAGTCTAA ATTTTTTGTA AAACTTAAAG TAAGAGTTTA      3300
Gly Asn Asn Thr Lys
245

TGCTTTGGCA TTGTTTGAAG TTCACTTACT AATGACTTTA G GCA GCA ATT ATG        3353
                                                Ala Ala Ile Met

AGG ATC GGG TTA CGG GAG ATG AAG GCA TTC TCC AAG CTT TTG TTT CCA      3401
Arg Ile Gly Leu Arg Glu Met Lys Ala Phe Ser Lys Leu Leu Phe Pro
250                 255                 260                 265

TCT GTT AAG  GAC ACT ACT TTC TTC GAG AGC TGT GGA GTT GCT GAC CTC    3449
Ser Val Lys  Asp Thr Thr Phe Phe Glu Ser Cys Gly Val Ala Asp Leu
                270                 275                 280

ATC ACA ACT TGT T GTAAGGAAGC ATATAGATTT CCTTCGAATA TGAATAAATT       3502
Ile Thr Thr Cys
285

GCATAGTTCA TATCATCATA ATTTGTGTTT GTGCTCAG TG GGC GGG AGA AAC        3554
                                            Leu Gly Gly Arg Asn
                                                        290

AGA AAA GTT GCT GAG GCT TTT GCA AAG AAT GGC GGG GAA AG              3595
Arg Lys Val Ala Glu Ala Phe Ala Lys Asn Gly Gly Glu Arg
            295                 300

GTCGTGTTTC CCTTTCGTCG ATCCTGATTT AATTCCTGTT TAGTGGTATT CACTTTGTGT   3655
GTATGTAAAT CAAGCAACTA TTTCCATCAT CTTCAG G TCA TTC GAT GAT CTC        3707
                                        Ser Phe Asp Asp Leu
                                        305

GAA GCA GAG CTG CTC CGG GGG CAA AAA TTA CAG GTACATGATG AAGAAACCGA   3760
Glu Ala Glu Leu Leu Arg Gly Gln Lys Leu Gln
310                 315                 320

TGTCTATACA GAAAGAGTCC ATTGCAAAGC TTGAGAATGT TTCGAGCATA AAGAGCATAA   3820

GAATATTCTT TTCGGTGATT TTCATGCAG GGT GTC TCA ACA GCA AAG GAG GTC      3873
                                Gly Val Ser Thr Ala Lys Glu Val
                                            325

TAT GAA GTC TTG GGG CAC CGA GGC TGG CTC GAG CTG TTC CCG CTC TTC      3921
Tyr Glu Val Leu Gly His Arg Gly Trp Leu Glu Leu Phe Pro Leu Phe
    330                 335                 340

TCA ACC GTG CAC GAG ATC TCC ACT GGC CGT CTG CAT CCT TCA GCC ATC      3969
Ser Thr Val His Glu Ile Ser Thr Gly Arg Leu His Pro Ser Ala Ile
345                 350                 355                 360

GTC GAA TAC AGC GAA CAA AAA ACC ATC TTC TCT TGG TAGAGCAAGA          4015
```

Val Glu Tyr Ser Glu Gln Lys Thr Ile Phe Ser Trp
365 370

GGCTGCCCTT GAAAGACTAA GAGCCACCCT GCCCTGTTTA AAGGGCTAAA AGTTTAATAT 4075
TTCTCTGCAG CCTAAACAGT TGGAAACATT GAAAATCTAG GATGTATCAG AAAAAAGAAG 4135
GTTTGGAGGA AGTATGGATG ATATAGAGGA CATGAATGTA TTCATTTTCG GTATACTCTT 4195
CTGCAAA ATAATTCTTC AGATGTTTTT GTGGTATGAG ATATAGAGGA CATGTATGTA 4255
TGCGGTAAGG CTGAAGTAAA CAAGTTACCA TAAGAGACAG CCCTCTCGGT TTCTTCCATC 4315
TGATCGATTC GTCTCGTCGA ATTTGCCAAA AGCTCAAAAC TCAACTCATC CCCTGCTTTC 4375
TATCCATATG GGCAAGGAAT ACAATTAGAC CAGTTTGATA CTTGTAATGA GAAGTTTAC 4434

(2) INFORMATION FOR ID SEQ NO:6

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 2955 base pairs

(B) TYPE: Nucleic acid

(C) STRANDEDNESS: Double strand

(D) TOPOLOGY: Linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: No

(iv) ANTI-SENSE: No

(vi) ORIGINAL SOURCE:

(A) ORGANISM: **Cuphea lanceolata**

(vii) IMMEDIATE SOURCE:

(A) LIBRARY: Genomic lambda FIX II

(B) CLONE: ClGPDHg3

(ix) FEATURE:

(A) NAME/KEY: CAAT signal

(B) LOCATION: 1055 to 1058

(ix) FEATURE:

(A) NAME/KEY: TATA signal

(B) LOCATION: 1103 to 1107

(ix) FEATURE:

(A) NAME/KEY: Start codon

(B) LOCATION: 1182 to 1184

(ix) FEATURE:

(A) NAME/KEY: CDS

(B) LOCATION: Join (1182 to 1326, 1837 to 1913, 2010 to 2082, 2180 to 2397, 2480 to 2587, 2668 to 2731, 2848 to 2885, 2947 to 2955)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:6

GGATCCTCCT CGATGGTGGT CCAATGAAGA CTATACAAAA CCAAGCCGAC GGAATCCGGT 60

GCACAATAAC TTGAAGCCAT GAAAACCAAT GCAATATATA GAGTACGCCT TGTACTATGT 120

AATATATTTA CAATTTTCTC TTGAATAGTT TAGGTTTGGT GATCGTAAAC TCGCAAAACA 180

CATATGTGCG TGTGTAA.ATA TATCTGGTGA TGATGTATGA AGAGAGTGCG GTTTAATTAC 240

CCGGTATTGT ATAAGGTTGT ATCTGCAGTT GACACTTTCA GTAGAAATTA CTAATAACTC 300

GACGAGATAC AAACGACTCG AGTTTCAGAA ATAAGTGGCA AAACGTTATG GGGTTCTCCT 360

TGATTCTTCG TGGAAGGTAT ACTATTAATC ATGTTCGCCT CCGTCCTAGT AGAAACATAG 420

AGTTTTTATC GGGATGCAGA TTGCAGATGA TAGAACTATT GTCAGATTCA TTATGCATAT 480

AGGATAGGCC TTCTACTGAT TTGGAAACTT ATATCGATTC TGTTGGAATG GATGTATGAA 540

AAGCTTCATA TCCGACATTG AAAATTTGGT CATATCAATA AGATGAACTA ACAAAATATG 600

CCAACCTCTT GGAAGCAAAA CACATCCGAG ACTTTAAGAT GTGGCTGAGG TTTCTGCAAC 660

TTTAAATCTC CCATATGCTT GACAGAATTG GTAGACCTAA CTCAATGGAT TTCATTCAAT 720

GATCGAAGTT TCTCTATCGA TCATAGCTGT GAATTAGTAA GCAAATGTCC ATAATATATC 780

CCCGAAAACA CGTAAAGTTA GGTCTCATTA CATTAGGCCT CAACCATATG TTATAAGTAA 840

ATTTGTTTTT TTTTTTTTCT CTTACAGTTG AATGTATCAA ATCGAAAAAA CCGTTAAGTC 900

GTTGCGGCCC TTTGAATAGT AAGCCAAAGA TCCGAAAGAA AAAGTAAACA GAGACAGAGC 960

AATGAGGAGA TGGCCAGTTT GAGAAGCAAA CGCATAGGTT GCCACGGAGG AGGCGGAGAC 1020

GGGTCATCGA TGACTTTCTC CGCCTCCTTA ACCGCAATGG CGATGCCGCC ATACCTCTCT 1080

```
GTCACCCTCT CTCCATTCCC TTTATATCTC TCCCGCTTCT TCCTCTGCTC CACTCAACCC   1140

CCTCTGCATA AACTCTGTGC TTTTTTAGTC TCTCCCCTGC T ATG TCG CCG GCA       1193
                                              Met Ser Pro Ala
                                                1

TTC GAA CCC CAT CAG CAG AAG CCT ACC ATG GAG AAC ATG CGA TTC CGA     1241
Phe Glu Pro His Gln Gln Lys Pro Thr Met Glu Asn Met Arg Phe Arg
  5                  10                  15                  20

GTC ACC ATC ATT G4GC AGC GGT AAC TGG GGC AGC GTC GCC GCT AAG CTC    1289
Val Thr Ile Ile Gly Ser Gly Asn Trp Gly Ser Val Ala Ala Lys Leu
                 25                  30                  35

ATT GCC TCC AAC ACC CTC AAC CTC CCG TCT TTC CAC G GTTTGTCTGC       1336
Ile Ala Ser Asn Thr Leu Asn Leu Pro Ser Phe His
             40                  45

CACTCTTCTT TCTTCATGAT CAGGCTCTTG CCAGTAGAGA CATGTCTTTT CATGAATCAA   1396

GCACCCGTTT TTTCGATGAG GATCACTGAG TTTGATTTAA GGGTATCCGA TGCAACTGCT   1456

GAAAAGATGT GGTTATTTTT GTTCTTTCAT GAAGTATCAT CTGAGAAATT TGATCTTAGC   1516

CTAAGCGGCA TTACTTTCGG TGTTAAGTTC ATTCTATGTG AGTAGGAGTA TGAGGTGATG   1576

CCGCGTGATT CCAATCAGGT ACCGATGAAA ATCAGTAGAC ATGGTTGCAG TTGAGGTTCC   1636

ATAGTTTACA CAGCATAGGA GTTGCTGTAT TTCTATTGAC GCTTGGATTT GTTTGGTGCT   1696

TATAATCCCG GTTTTTACTA ATTGGTTATG AACACCGATA ATAACAACAG TTAGATTTCT   1756

TCAACATTAA CCGGTTGAAG ATTAGGCCAT ATTCTTATTT GGGTACTATT TCTTAAGAAA   1816

ACATTCATAT TTTCTTTCAG AT GAA GTA AGG ATG TGG GTG TTT GAG GAG         1865
                      Asp Glu Val Arg Met Trp Val Phe Glu Glu
                           50                  55

ACA TTG CCA AGC GGC GAG AAG CTC ACT GAA GTC ATC AAC CGG ACC AAT     1913
Thr Leu Pro Ser Gly Glu Lys Leu Thr Glu Val Ile Asn Arg Thr Asn
     60                  65                  70

GTAAGGAAGA TCAATTTAGC ATGTCATTGT ATTAACATAA AGAGCGTTTA TTGGCAACTT   1973

TGGCTTTCAT GATGTTCGAG TGTTGCGTCT TTGCAG GAA AAT GTT AAG TAT CTG     2027
                                       Glu Asn Val Lys Tyr Leu
                                        75                  80

CCT GGA TTC AAG CTT GGC AGA AAT GTT ATT GCA GAC CCA AAC CTT GAA     2075
Pro Gly Phe Lys Leu Gly Arg Asn Val Ile Ala Asp Pro Asn Leu Glu
                 85                  90                  95
```

```
AAT GCA G GTAGTGATTG TATTTCAGTG CTCGGTTGAA TGATCAAGTA AAATCCTCGT      2132
Asn Ala

GCTAAATATG TCGAGATGTT CGTGTTTTTG CATAATGTTT TGTTTAG TT AAG GAA        2187
                                                    Val Lys Glu
                                                        100

GCA AAC ATG CTT GTA TTT GTC ACA CCG CAT CAG TTC GTG GAG GGC CTT       2235
Ala Asn Met Leu Val Phe Val Thr Pro His Gln Phe Val Glu Gly Leu
            105                 110                 115

TGC AAG AGA CTC GTC GGG AAG ATA AAG GCA GGTGCA GAG GCT CTC TCC        2283
Cys Lys Arg Leu Val Gly Lys Ile Lys Ala Gly Ala Glu Ala Leu Ser
        120                 125                 130

CTT ATA AAG GGC ATG GAG GTC AAA AGG GAA GGG CCT TCC ATG ATA TCT       2331
Leu Ile Lys Gly Met Glu Val Lys Arg Glu Gly Pro Ser Met Ile Ser
    135                 140                 145

ACC TTA ATC TCG AGC CTT CTC GGG ATC AAC TGC TGT GTC CTA ATG GGA       2379
Thr Leu Ile Ser Ser Leu Leu Gly Ile Asn Cys Cys Val Leu Met Gly
150                 155                 160                 165

GCA AAC ATC GCC AAC GAG GTAAAATCTT GGTGCAGTCT TACGAGATTC              2427
Ala Asn Ile Ala Asn Glu
                170

TGAATCTTGA ACCTGTTAGC ATTTTGACAC ACTGTGACTT CTAAATTTGT AG ATT         2482
                                                            Ile

GCT CTT GAG AAA TTC AGT GAG GCG ACA GTC GGA TAC AGA GAA AAT AAG       2530
Ala Leu Glu Lys Phe Ser Glu Ala Thr Val Gly Tyr Arg Glu Asn Lys
        175                 150                 185

GAT ACT GCA GAG AAA TGG GTT CGG CTC TTC AAC ACT CCA TAC TTC CAA       2578
Asp Thr Ala Glu Lys Trp Val Arg Leu Phe Asn Thr Pro Tyr Phe Gln
    190                 195                 200

GTC TCG TCT GTGAGTACGA ATAAACCTTT CCTTCTGCGA ACAAAAAACT              2627
Val Ser Ser
205

TCCCGAGGCA GGAACTAAAT GAAACAAGTT AACATAATAG GTT CAA GAT GTG GAA       2682
                                            Val Gln Asp Val Glu
                                                210

GGA GTG GAA CTT TGT GGC ACA CTG AAG AAT GTC GTG GCC ATA GCA GCC G     2731
Gly Val Glu Leu Cys Gly Thr Leu Lys Asn Val Val Ala Ile Ala Ala
        215                 220                 225

GTACTTATAT ACGATCTCCA CATTTATATA AACTAGTTAG AAAGATTTTG GATTGCTGTA    2791

AAAACCGTGG AAAAACCCGA AAAGTGTTGA TGAAGTGTTA CCAAATGTTG TTTCAG GT     2849
                                                            Gly
```

```
TTT GTA GAT GGA CTG GAG ATG GGA AAC AAC ACA AAG GTAAGTCCAA          2895
Phe Val Asp Gly Leu Glu Met Gly Asn Asn Thr Lys
230                 235                 240

AGTTCATGCA AATTTTTTCG TATTTACGAC TGAATGCTTG GATATACATA G GCT GCG    2952
                                                          Ala Ala

ATT                                                                2955
Ile
```

(2) INFORMATION FOR ID SEQ NO:7

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 574 base pairs

(B) TYPE: Nucleic acid

(C) STRANDEDNESS: Double strand

(D) TOPOLOGY: Linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: No

(iv) ANTI-SENSE: No

(vi) ORIGINAL SOURCE:

(A) ORGANISM: Cuphea lanceolata

(vii) IMMEDIATE SOURCE:

(A) LIBRARY: Genomic lambda FIX II

(B) CLONE: ClGPDHg3

(ix) FEATURE:

(A) NAME/KEY: CDS

(B) LOCATION: 31 to 189

(ix) FEATURE:

(A) NAME/KEY: Stop codon

(B) LOCATION: 190 to 192

(ix) FEATURE:

(A) NAME/KEY: PolyA signal

(B) LOCATION: 393 to 398

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:7

```
GGCATATCGA TGATTTTTCC TATCTTGCAG GGT GTC TTG ACA GCA AAA GAG GTG      54
                                Gly Val Leu Thr Ala Lys Glu Val
                                  1               5

TAT GAG GTA CTG AAG CAC CGG GGC TGG CTC GAG CGT TTC CCG CTC TTC      102
Tyr Glu Val Leu Lys His Arg Gly Trp Leu Glu Arg Phe Pro Leu Phe
     10                  15                  20

GCA ACT GTG CAT GAG ATC TCA TCT GGC AGG TTG CCT CCT TCA GCC ATT      150
Ala Thr Val His Glu Ile Ser Ser Gly Arg Leu Pro Pro Ser Ala Ile
 25                  30                  35                  40

GTC AAA TAC AGC GA-A CAA AAG CCC GTC TTA TCT CGA GGT TAGAACGAGA      199
Val Lys Tyr Ser Glu Gln Lys Pro Val Leu Ser Arg Gly
                45                  50

GAAAACCCGA CAAACCGGTG AAACTCGTAG TCTTAAACTG AAATCCAAAA ACATGCTGGG    259
AACATCAGCA AAAACCATTC ATCAAGGATG TCTTAGATAA AAGGTTTCAG GAAGAAATAG    319
ATGGTAGTGT GTGTAATGTT ATCAGCAATC ATTCATTCAT TTATTAAGTA TTTTTTGCAT    379
CATATTTTAT GCTAATAATT ATTACATAAA TTACTCAAAT TTTGTCAAAA TTTCTGCATT    439
GCCCCAAACA GATTAATGCA TTGAGAAAAA CTTATAAAGC TTTATCCAGC ATACATATAG    499
TTCTTTAAGC AATACAAAAA CACCCTTCTA AGCCTCTTTG AAGATGGAGT TTGATCACAC    559
ATTAAAATGC TTTTT                                                    574
```

(2) INFORMATION FOR ID SEQ NO:8

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 1507 base pairs

(B) TYPE: Nucleic acid

(C) STRANDEDNESS: Double strand

(D) TOPOLOGY: Linear

(ii) MOLECULE TYPE: DNA (molecular)

(iii) HYPOTHETICAL: No

(iv) ANTI-SENSE: No

(vi) ORIGINAL SOURCE:

(A) ORGANISM: Cuphea lanceolata

(vii) IMMEDIATE SOURCE:

(A) LIBRARY: Genomic lambda FIX II

(B) CLONE:ClGPDHg9

(ix) FEATURE:

(A) NAME/KEY: TATA signal

(B) LOCATION: 1108 to 1112

(ix) FEATURE:

(A) NAME/KEY: Start codon

(B) LOCATION: 1193 to 1193

(ix) FEATURE:

(A) NAME/KEY: CDS

(B) LOCATION: 1193 to 1376

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:8

GCATGCGGGC AGGCAGGCAG GCATGGGTCT AAATTCTAGA AGACCCAGAC ATATTCATTT 60

TGTTCACAAC CGACCCATCA ATATATTGAT TAATTTTGTT TAAATTTATC ATCAGTTTTT 120

ATTTAATATT TTTAAATAGG TTTACCTTGA TCGTGATAAT TATTTAATAT TACTTTGTAA 180

TAGTTTATTT ATCTAGCGTT ATAAAATAAC ATTTGAATTC GTTGATGATA TGTGTATTTT 240

TACTATGTTT ATATGAAATT TATATTTCAA ATATTAAATA ATGTTCTTAT TTTGGCCTAT 300

GGAGAAGTAT CATCAATTTT TCTATTAAAT AACAGTCTTC AGTTTAGTCA AATCAGTTGA 360

TAAGTTCCCA AATCACACAT TGTTTGTATG AAAATTTTAA TAAAAAAGTT AAGATGGTAT 420

TATTATAGAA AAATATATAA AGTATCTTTA AATAATAATT TCTTTTTAAT ACAAAAAGGA 480

ATATTTGATT ACTTGACTTA TAAAATTTAT TGATAAGGAT GCCAACTTTC ATTTTAGAAA 540

CTAGAGTAAT GATGGTTAAA TTCCCCGAAA AATGGTATGT CAATTTATTG ATACGTTCCA 600

CTACTATTT CTGAGACATT TACATGTTTG TAAAAAAAAT CTATATATTT AAATTAAGAT 660

GGGTGTAATC AATTATAAAA TACAGCGAAT TTTAACACCG AATGAATAGA TTATCTGCAT 720

AACAATTTAT ACCATCCCTA AATACGAATT AGCAAGTTAA TAAAATTTAA TTACACGAAC 780

```
CATGATTATA TAAATTATCG AATCCCCGAC GTGGGGACGT ACCGAACCAA CCGTTGAAGT     840

GGTTGCCCTT TGAATCCTAA GACATACAGA CGTCATGATT CTTTGTCTCT CTATCTGTCC     900

ATTTACATAA TAAAATCAAA GAGAAGAAAA CAGAGGAAGC AGAGCATAGC ATAGCATAGC     960

ATAGAGGAGA TCGCCAGATT CAGCTGTTTC CTCATAGTTT GCCACGAGAC ATACATTGCA    1020

TTGCCCGATG CCTTTCTCCG CCTCCTTGTC CCTCTCCTCA TTCCCCCGAT GCCTTTCTCC    1080

GCCTCCTTGT CCCTCTCCTC ATTCCCTTAT ATCCCTCCTC CCCTCCCTCT TCTTCCTCTG    1140

CTCAACTCCT CCCCCTCACC CTCTTCCTCT OTTCTTCCTC TCTGCCTCTG CA ATG        1195
                                                          Met
                                                           1

GCG CCT GCC TTC GAA CCC CAT CAG CTG GTT CCT TCT GAG CTT AAC TCT       1243
Ala Pro Ala Phe Glu Pro His Gln Leu Val Pro Ser Glu Leu Asn Ser
             5                  10                  15

GCC CAC CAG AAC CCA CAT TCC AGC GGA TAT GAA GGA CCC AGA TCG AGG       1291
Ala His Gln Asn Pro His Ser Ser Gly Tyr Glu Gly Pro Arg Ser Arg
        20                  25                  30

GTC ACC GTC GTT GGC AGC GGC AAC TGG GG4C AGC GTC GCT GCC AAG CTC      1339
Val Thr Val Val Gly Ser Gly Asn Trp Gly Ser Val Ala Ala Lys Leu
    35                  40                  45

ATT GCT TCC AAC ACC CTC AAG CTC CCA TCT TTC CAT G GTTAGTCTCT          1386
Ile Ala Ser Asn Thr Leu Lys Leu Pro Ser Phe His
 50                  55                  60

CATTCTTCTC TCTGTAAAGT TGAAGCTTTT TCATGGAATA GTCTCTAGAC ATGAGCCCCT    1446
GTTTGCATGG TTTTGTTTTG TCTTTGAAAC ATGAATAAAG GTGGTTTCTT GTGTTGGTAC    1506
C                                                                    1507
```

Patent Claims

1. DNA sequences which are isolated from plants and code for a glycerol-3-phosphate dehydrogenase, and the alleles as well as derivatives of these DNA sequences.

2. DNA sequences according to claim 1, wherein they are isolated from **Cuphea lanceolata**.

3. Genomic clones which are isolated from genomic plant DNA and contain a complete gene of a glycerol-3-phosphate dehydrogenase and the alleles as well as derivatives of this gene.

4. Genomic clones according to claim 3, wherein the complete gene contains the promoter sequence and other regulator elements in addition to the structure gene.

5. Genomic clones according to claim 4, wherein the plant DNA originated from **Cuphea lanceolata**.

6. Promoters and other regulator elements of the glycerol-3-phosphate gene from one of the genomic clones according to claims 3 to 5, and the alleles as well as the derivatives of these promoters.

7. DNA sequences according to claim 1, obtained from functional complementation with mutants of a microorganism.

8. DNA sequences according to claim 7, wherein the microorganism is E. coli BB26-36.

9. Procedure for producing plants, plant parts and plant products the triacylglyceride content or fatty acid pattern of which is altered, in connection with which a DNA sequence is transferred according to one of

claims 1 or 2, or a gene originating from the genomic clones according to one of claims 3 to 5 is transferred by genetic engineering methods.

10. Procedure according to claim 9, wherein the DNA sequence or the gene is transferred by microinjection, electroporation, particle gun, steeping of plant parts in DNA solutions, pollen or pollen tube transformation, transfer of corresponding recombinant Ti plasmids or Ri plasmids with **Agrobacterium tumefaciens**, liposome-mediated transfer, or by plant viruses.

11. Use of a DNA sequence according to one of claims 1 or 2 or of a gene originating from the genomic clones according to one of claims 3 to 5 for altering the biosynthesis output in plants.

12. Plants, plant parts and plant products produced according to a procedure of claims 9 or 10.

AMENDED PAGE

IPEA/EP